IEICE Electronics Express, Vol.5, No.17, 650–656

An efficient hybrid distributed video coding

Anhong Wang[1,2a)], **Yao Zhao**[1], **and Jeng-Shyang Pan**[3]

[1] *Institute of Information Science, Beijing Jiaotong University, P.R. China*

[2] *Taiyuan University of Science and Technology, P.R. China*

[3] *Kaohsiung University of Applied Sciences, Taiwan*

a) *wah_ty@yahoo.com.cn*

Abstract: An efficient hybrid distributed video coding (HDVC) scheme combining residual coding, SW-SPIHT coding and intra-mode decision technology is proposed to offer good rate-distortion performance and low-complexity encoding. The scheme adopts the wavelet-domain Wyner-Ziv codec to compress the residual frame between the current frame and its reference frame. An intra-mode decision based on the temporal and spatial correlation is made to determine whether a wavelet block is better coded or not. When the estimated correlation is weak, we use intra-SPIHT coding; otherwise SW-SPIHT (Slepian-Wolf SPIHT) coding is adopted. Experimental results indicate the proposed HDVC achieves evident improvement than the referenced.

Keywords: DVC, residual coding, SW-SPIHT, intra-mode decision

Classification: Science and engineering for electronics

References

[1] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," *Proc. Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE, Special Issue on Video Coding and Delivery*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[3] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv residual coding of video," *Proc. Picture Coding Symposium*, Beijing, China, April 2006.

[4] M. Tagliasacchi, A. Trapanese, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Intra mode decision based on spatio-temporal cues in pixel domain Wyner-Ziv video coding," *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[5] D. Tsai, C. Lee, and W. Lie, "Dynamic key block decision with spatio-temporal analysis for Wyner-Ziv video coding," *Proc. IEEE International Conference on Image Processing*, San Antonia, USA, Sept. 2007.

[6] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *Proc. Asilomar Conference on Signals, Systems and Computers, Pacific Grove*, CA, Nov. 2005.

[7] A. Wang, Y. Zhao, Z. Zhu, and H. Wang, "Scalable distributed video coding based on block SW-SPIHT," *Chinese Optics Letter*, vol. 5, no. 6, pp. 336–339, June 2007.

© IEICE 2008

DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

1 Introduction

Distributed video coding (DVC) [1, 2] is a new video coding framework based on 1970's Slepian-Wolf and Wyner-Ziv theories. The main goal of DVC is to achieve low-complexity encoding to meet the demands of friendly uplink communication services, possibly at the expense of high-complexity decoding. Generally, the current DVC schemes are generalized as two approaches, including pixel-domain DVC and transform-domain DVC. Compared to the former, the latter uses a transform-domain Wyner-Ziv codec instead of the pixel-domain one. For example, some transform-domain Wyner-Ziv codecs [1, 2] applied discrete cosine transform (DCT) to exploit spatial correlation within a frame, but the selection of the combined quantization for DCT coefficients bands is an arduous task. Anne Aaron [3] proposed a residual DVC to compress the residual frame between the current frame and a reference frame through pixel-domain Wyner-Ziv codec. Due to the exploitation of the temporal correlation between the current frame and the reference frame at the encoder, residual DVC achieves almost the same rate-distortion performance as DCT-domain DVC. Meanwhile, the residual DVC overcomes the difficulty of the combined quantization selection. However, there is still some spatial correlation remained in residual frames in [3]. The rate-distortion performance should be enhanced further if a proper transform coding with easily combined quantization methods is adopted to the residual frame. In this paper, a wavelet-domain Wyner-Ziv codec with SPIHT (set partitioning in hierarchical trees) is applied to the residual frame, which exploits the temporal and spatial correlation at the encoder simultaneously. Meanwhile, SPIHT brings convenience for the combined quantization of wavelet coefficients.

On the other hand, the current transform-domain Wyner-Ziv codec assumes the temporal correlation between the source to be encoded and its side information to be spatially stationary. In other words, different temporal correlation on the spatial location is neglectable. Though the block-based intra-mode decision in pixel domain [4, 5] is proposed to classify the temporal correlation on spatial location, there is still much need to exploit intra-mode decision to the transform-domain. In our work, a mode decision based on temporal and spatial correlation is also proposed. When the estimated temporal correlation on the spatial location is weak, we use intra-SPIHT coding on the corresponding wavelet block (WB); otherwise the inter mode coder SW-SPIHT is adopted.

Because of the combination of residual coding, SW-SPIHT coding and the intra-mode decision technique, the proposed DVC is named as hybrid DVC (HDVC). The experimental results show HDVC achieves up to 4.1 dB improvement in PSNR than pure SW-SPIHT, 3.2 dB improvement than residual DVC, and 3 dB than the latest DCT-domain DVC of DISCOVER [1] with some neglectable complexity. Besides, the intra-mode decision offers up to 2.1 dB improvement at high motion area.

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

2 The proposed HDVC framework

The framework of the proposed HDVC is shown in Fig. 1, compared to the current DVC schemes, our contributions mainly focus on: 1) wavelet-domain Wyner-Ziv codec is applied on WBs instead of frames to provide some conveniences for temporal and spatial correlation analysis; 2) instead of original frames, the residual frame is used to construct WBs and analyzed to exploit the temporal and spatial correlation; 3) an intra-mode decision of wavelet domain is conducted to each WB to exploit different temporal and spatial correlation; 4) the intra-mode decision strategy based on temporal and spatial criteria is given.

2.1 Wavelet-domain Wyner-Ziv codec design for residual frames

Residual DVC had been proven to have the performance similar to the DCT-domain DVC due to the exploitation of temporal correlation at the encoder [3]. Specially, the frame W_{re}, the simple estimation to the current frame W, acts as the reference frame at both encoder and decoder. However, there is still some spatial correlation in the residual frame, so we propose a wavelet-domain Wyner-Ziv to the residual frames to extract the temporal and spatial correlation simultaneously at the encoder. The reference frame W_{re} is obtained by the weighted average (WA) interpolation:

$$W_{re} = \alpha K'_j + \beta K'_{j+1} \tag{1}$$

where K'_j and K'_{j+1} are the previous and next decoded key frames; $\alpha = 1 - \ell/g$, $\beta = 1 - \alpha$, where ℓ is the number of frames between the current frame and the previous key frame K'_j in the same group of picture (GOP), g is the number of frames in a GOP.

At the encoder, the residual frame $D = W - W_{re}$ is decomposed by discrete wavelet transform (DWT) and the wavelet coefficients C_d are outputted.

At the decoder, the residual frame $D_y = W - W_{re}$ is decomposed by the same DWT and the resulting coefficients C_d^y are generated. Here, Y

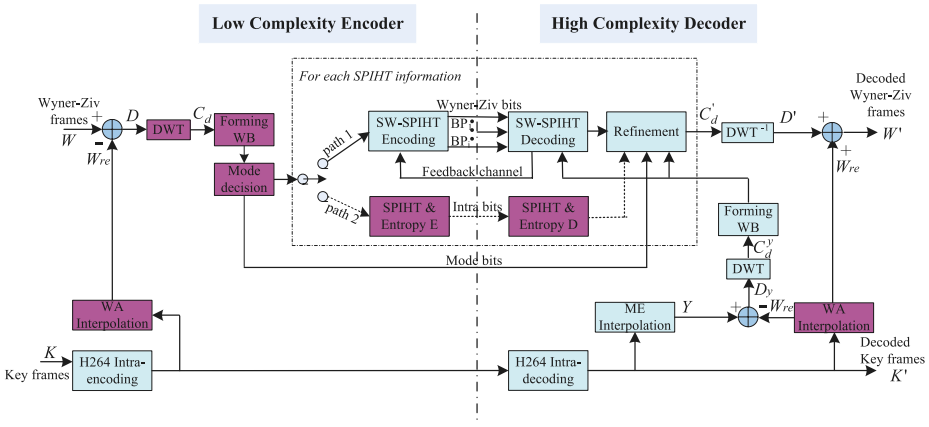


Fig. 1. The proposed HDVC framework

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

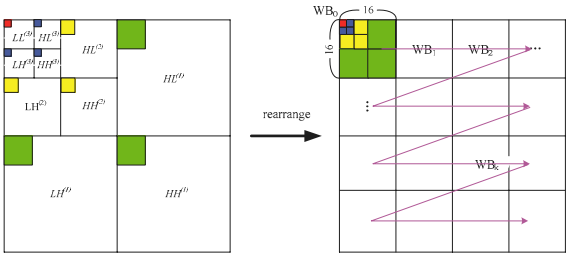


Fig. 2. WBs are rearranged and scanned in row-by-row

is the more accurate estimation to the current frame and is generated by motion-estimated (ME) interpolation [1, 3] of key frames.

2.2 Intra-mode decision strategy for WBs

In this paper, WB is the basic coding unit. Typically, C_d and C_d^y are partitioned into 16×16 cross-scale WBs as in Fig. 2. In each WB, there are four 2×2 coefficients vectors from $LL^{(3)}$, $HL^{(3)}$, $LH^{(3)}$ and $HH^{(3)}$, three 4×4 from $HL^{(2)}$, $LH^{(2)}$, and $HH^{(2)}$, three 8×8 from $HL^{(1)}$, $LH^{(1)}$ and $HH^{(1)}$. Obviously, a WB composes of several wavelet trees, so SPIHT is conveniently applicable in this case.

In Fig. 1, there are two encodings, intra-SPIHT and SW-SPIHT, to be determined for each WB. The encoding selection of each WB depends on the correlation derived from the temporal and spatial analyses. As a low encoding complexity scenario is targeted, a strategy which is easily implemented and capable to infer the correlation between the current WB and its side information is needed. Here, the WB co-located at C_d^y is used as the side information for the current WB at C_d.

Based on the low frequency energy E_{LL} of current WB to be encoded, the temporal criterion is calculated as:

$$E_{LL} = \sum_{i=1}^{N_{LL}} (C_i^{LL})^2 \tag{2}$$

where C_i^{LL} is the $i-th$ coefficient in $LL^{(3)}$ band, N_{LL} is the total number of coefficients in $LL^{(3)}$, $N_{LL} = 4$.

Meanwhile, based on the variance of high frequency coefficients of current WB, the spatial criterion is calculated as:

$$\delta^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (C_i^h)^2 - \left(\frac{1}{N_h} \sum_{i=1}^{N_h} C_i^h \right)^2 \tag{3}$$

where C_i^h is the $i-th$ coefficient in all high frequency bands, N_h is the total number of coefficients in high frequency bands, $N_h = 12$.

There are two cases that the intra-SPIHT is used to a WB. In the first case of $E_{LL} \geq T_1$, the temporal correlation between current WB and its side information is assumed weak. That is because E_{LL} collects most of the energy of WB, and the greater E_{LL} means the greater difference between the current frame W and the referenced frame W_{re}. While W_{re} is interpolated

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

by the key frames with zero-motion vector. Then, this is likely that the correlation with Y interpolated by key frames with non-zero motion vectors at the decoder is weak. So the correlation between current WB and its side information is weak. The second case occurs when $T_2 \leq E_{LL} \leq T_1$ and $\delta^2 \leq T_3$. In this case the low frequency coefficients collect most of energy and the high frequency coefficients are spatially concentrated on small values so the intra-SPIHT coding is suited. Otherwise in the case of $E_{LL} < T_2$, the temporal correlation is inferred as strong so SW-SPIHT is used. Here, T_1, T_2 and T_3 are the pre-determined thresholds by experiments.

At the encoder, according to the intra-mode decision, two kinds of bit streams, the Wyner-Ziv bits and the intra bits, are formed respectively by path 1 and path 2 as depicted in Fig. 1. The Wyner-Ziv bits consist of the syndrome bits of SW-SPIHT encoding. The intra bits refer to the SPIHT information bits compressed by entropy coding. In our work, different SPIHT information streams in each bit-plane are encoded respectively. In other word, the tree distribution information, the significant information, the refinement information, and the sign information are formed respectively by a row-by-row scanning among the blocks with the same mode. Each SPIHT information stream is compressed either by entropy coding or by SW coding with low-density-parity-check-accumulate (LDPCA) [6]. LDPCA with 396 nodes is adopted in our work. In the case of the sequence length less than 396, the padding zeros are applied. The details of SW-SPIHT for WBs are referred to the literature [7]. The binary mode bits are also encoded with entropy and sent to the decoder.

At the decoder, the bit streams are recovered either by LDPCA decoding or entropy decoding. The decoding selection depends on the decoded mode bits. In addition, refinement process effectively reconstructs the current WB with side information WB at C_d^y. Next, the recovered coefficients C'_d are arranged and the residual frame D' is generated by inversed DWT. Finally, we have the recovered pixels from the residual frame and the reference frame as follows

$$W' = D' + W_{re}. \tag{4}$$

3 Experimental results

Experiments are conducted to the luminance component of standard QCIF@15 fps (frame per second) sequences. Five different DVC schemes are compared, including HDVC with intra-mode proposed, HDVC without intra-mode, pure SW-SPIHT DVC, the residual DVC and the DCT-domain DVC. Fig. 3 and Fig. 4 show the results.

First, compared with the pure SW-SPIHT results computed based on [7], the proposed HDVC obtains up to 4.1 dB for hall sequence. While compared with residual DVC results which are computed based on paper [3], the improvement is up to 3.2 dB. These improvements originate from the fact that the hybrid scheme effectively employs some temporal and special correlation simultaneously at the encoder. Besides, compared with the latest DCT-

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

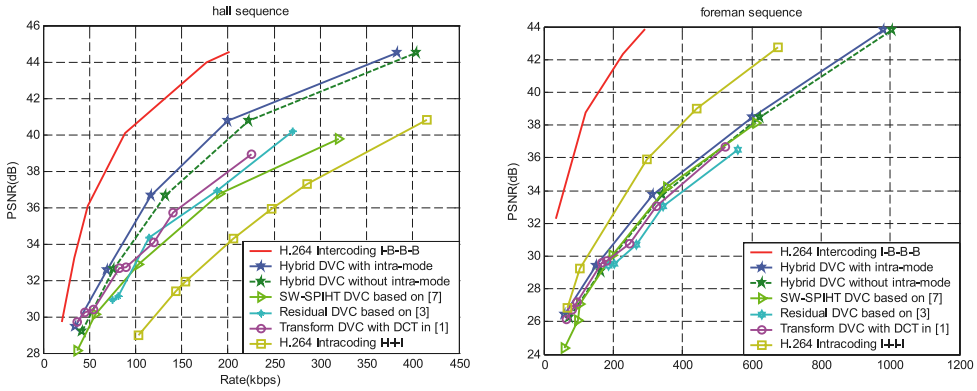


Fig. 3. Rate-distortion comparison when GOP = 8

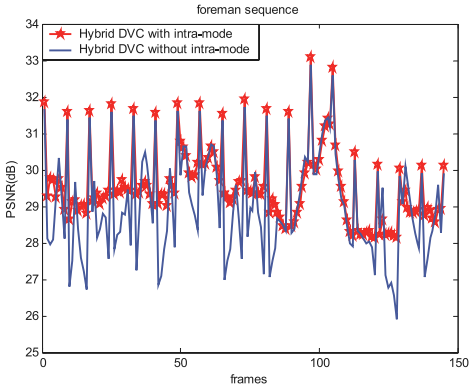


Fig. 4. Frames quality comparison when GOP = 8 and bit-rate = 166 kbps

domain DVC [1] results obtained from DISCOVER (www.discoverdvc.org), the proposed HDVC obtains up to 3 dB improvement for hall sequence. For high motion sequence foreman, there is only up to 0.9 dB improvement because the temporal correlation is weak.

Second, with the intra-mode decision, the evident quality improvement is achieved especially in the high motion frames. In Fig. 4, the intra-mode decision contributes up to 2.1 dB improvement in PSNR in high-motion case.

In addition, compared with intra-frame coding of H.264/AVC, the proposed scheme has up to 7 dB enhancement in PSNR. However, there is still a gap from the state-of-the-art inter-frame coding of H.264/AVC.

4 Conclusion

In this paper, an efficient hybrid DVC with intra-mode decision is presented to improve the rate-distortion performance of current DVC. This scheme benefits from the combination of residual coding, SW-SPIHT coding and intra-mode decision strategy. It achieves evident improvement than the pure SW-SPIHT, residual DVC and the latest DCT-domain DVC of DISCOVER. The encoding of HDVC mainly consists of SPIHT, intra-mode decision, and LDPCA encoding. Thus, the proposed scheme avoids the complex motion estimation and preserves the low-complexity encoding similar to the intra-

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008

frame coding. While, compared to the referenced DVC, our HDVC has a little more complex encoding due to the combining of mode decision and the subtraction operation though they are neglectable. The proposed scheme is promising in the low-complexity encoding applications such as mobile cameras, wireless low-power surveillances.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 60776794, No. 90604032), 973 program (No. 2006CB303104), 863 program (No. 2007AA01Z175), PCSIRT (No. IRT0707) and Youth Science Technology Research Foundation of Shanxi Province (No. 2008021020).

© IEICE 2008
DOI: 10.1587/elex.5.650
Received June 08, 2008
Accepted July 16, 2008
Published September 10, 2008